INSTAGRAM USER ANALYTICS

By Vaibhav

Project Description

Instagram User analysis involves tracking how users engage with a digital product, such as a software application or a mobile app. The insights derived from this analysis can be used by various teams within the business. For example, the marketing team might use these insights to launch a new campaign, the product team might use them to decide on new features to build, and the development team might use them to improve the overall user experience.

In this project, I used SQL and MySQL Workbench to analyze Instagram user data and answer questions posed by the management team. My insights will help the product manager and the rest of the team make informed decisions about the future direction of the Instagram app.

The goal of this project is to use SQL skills to extract meaningful insights from the data. My findings could potentially influence the future development of one of the world's most popular social media platforms.

The project "Instagram User Analytics" is an extensive data analysis initiative aimed at extracting meaningful insights from an Instagram-like database for the purposes of marketing strategy, user engagement, investor relations, and overall platform health. This project provided critical data-driven findings that could help in decision-making processes across multiple strategic areas including user retention, campaign planning, user authenticity verification, and more.

Objective of this Project

A) Marketing Analysis:

Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.

1. **To identify the five oldest users on Instagram from the provided database.**

Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.

2. **To identify users who have never posted a single photo on Instagram.**

Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.

3. **To determine the winner of the contest and provide their details to the team.**

Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

4. **To identify and suggest the top five most commonly used hashtags on the platform**.

Ad Campaign Launch: The team wants to know the best day of the week to launch ads.

5. **To determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.**

B) Investor Metrics:

User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

1. **Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.**

Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.

2. **To identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.**

Approach

The project involved using structured query language (SQL) to analyze a relational database with tables containing information about users, photos, likes, and tags. Key focus areas included:

1. Loyal User Reward Program: Identifying the oldest users on the platform to potentially reward loyalty.
2. Inactive User Engagement: Finding users who never posted to re-engage them through promotional emails.
3. Contest Management: Determining the user with the most likes on a single photo to declare a contest winner.
4. Hashtag Research: Analyzing the most commonly used hashtags for strategic brand collaborations.
5. Ad Campaign Launch Timing: Discovering the most popular day of user registration to optimize ad campaign launches.
6. User Engagement Metrics for Investors: Calculating the average number of posts per user to provide an understanding of overall platform activity.
7. Detection of Bot and Fake Accounts: Identifying users who liked every post, which could indicate non-human behavior.

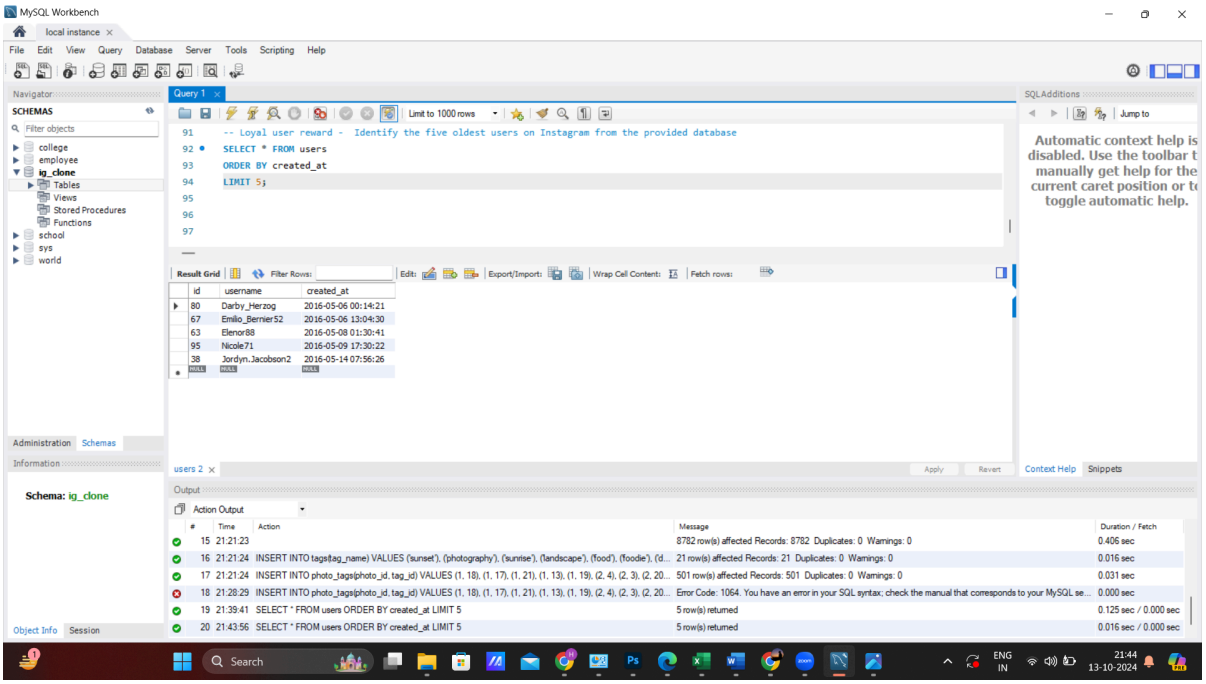
Tech Stack used

SQL: Used for querying and extracting data insights from the relational database.
Database Management System (DBMS): Likely employed for hosting and managing the user, photos, and interactions tables.

Marketing Analysis

1. Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.
Identify the five oldest users on Instagram from the provided database

```
SELECT * FROM users
ORDER BY created_at
LIMIT 5;
```



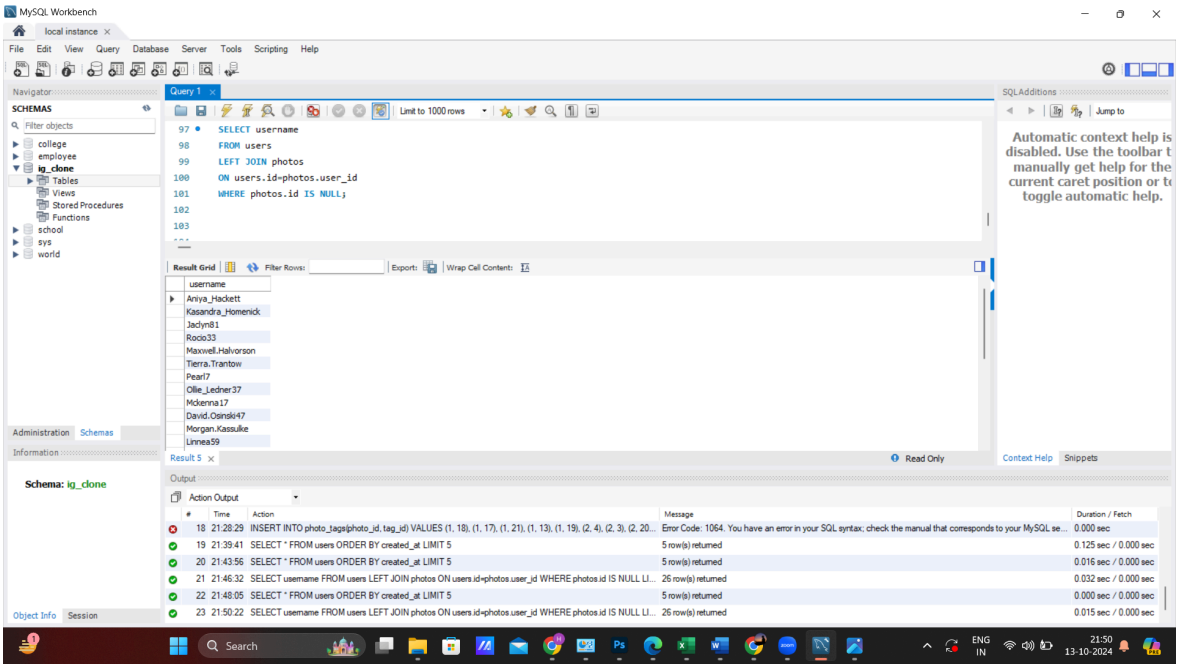
Result

I'd	Username.	Created at
80	Darby_Herzog	2016-05-06 00:14:21

67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
95	Nicole71	2016-05-09 17:30:22
38	Jordyn.Jacobson	2016-05-14 07:56:26

2. Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.
Identify users who have never posted a single photo on Instagram.

```
SELECT username
FROM users
LEFT JOIN photos
ON users.id=photos.user_id
WHERE photos.id IS NULL;
```



Output

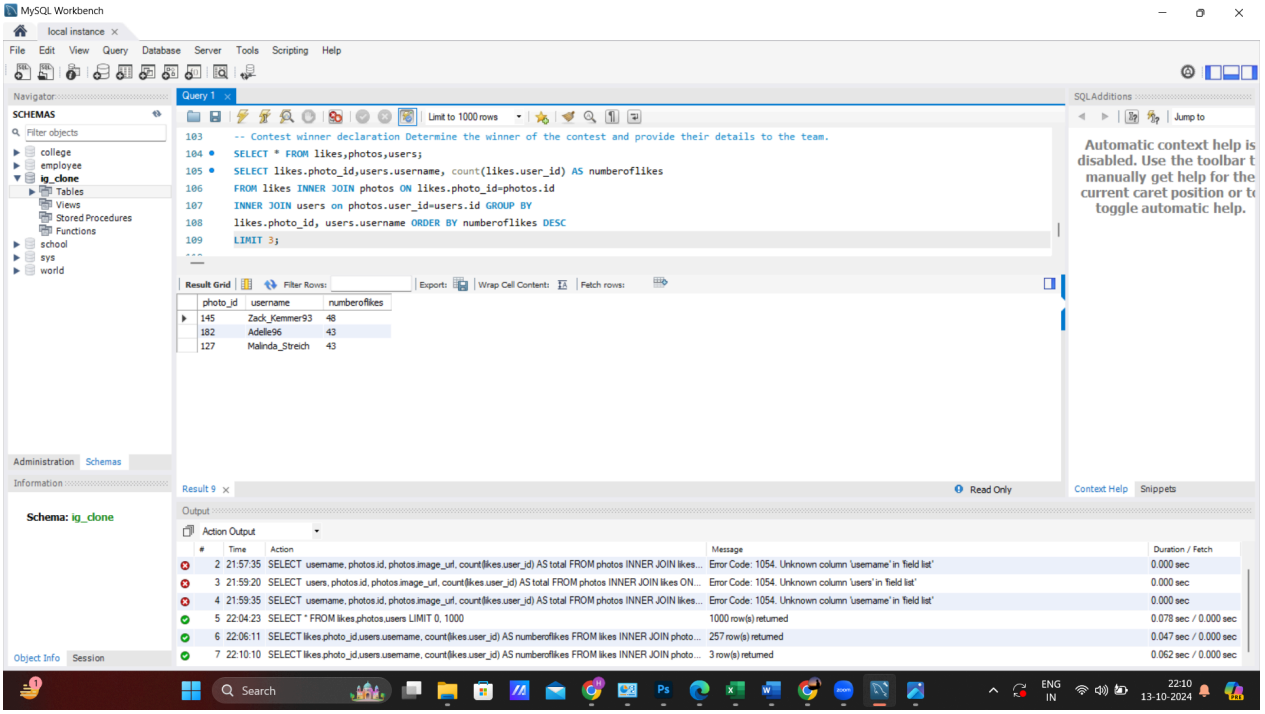
Here's the list of Instagram users who have never posted a single photo

Aniya_Hackett
Kasandra_Homenick
Jaclyn81
Rocio33

Maxwell.Halvorson
Tierra.Trantow
Pearl7
Ollie_Ledner37
Mckenna17
David.Osinski47
Morgan.Kassulke
Linnea59
Duane60
Julien_Schmidt
Mike.Auer39
Franco_Keebler64
Nia_Haag
Hulda.Macejkovic
Leslie67
Janelle.Nikolaus81
Darby_Herzog
Esther.Zulauf61
Bartholome.Bernhard
Jessyca_West
Esmeralda.Mraz57
Bethany20

3. Contest winner declaration: The team has organized a contest where the user with the most likes on a single photo wins.
Determine the winner of the contest and provide their details to the team.

```
SELECT * FROM likes,photos,users;
SELECT likes.photo_id,users.username, count(likes.user_id) AS numberoflikes
FROM likes INNER JOIN photos ON likes.photo_id=photos.id
INNER JOIN users on photos.user_id=users.id GROUP BY
likes.photo_id, users.username ORDER BY numberoflikes DESC;
LIMIT 3;
```



Output

Photo I'd.	Username.	Number of likes
145	Zack_Kemmer93	48
182	Adelle96	43
127	Malinda_Streich	43

Zack_Kemmer93 is the winner with the most number of likes; 48.

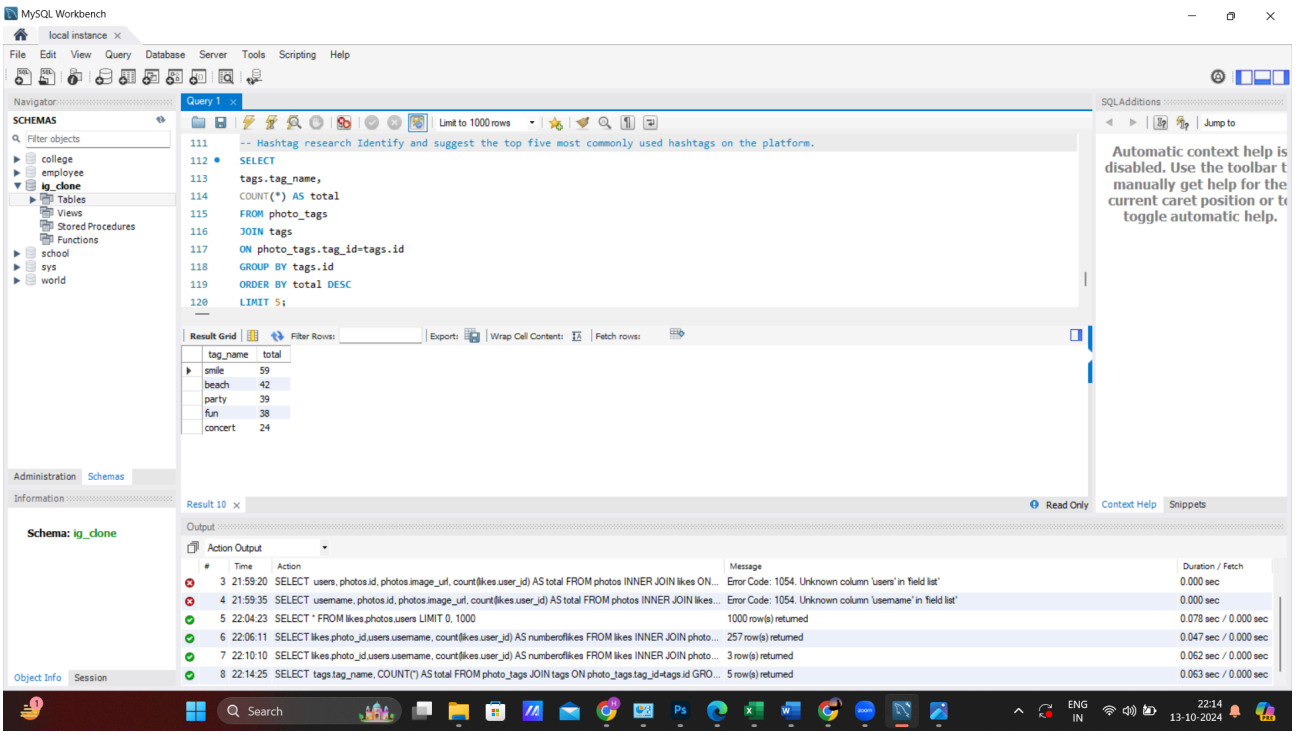
4. Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.
Identify and suggest the top five most commonly used hashtags on the platform.

```
SELECT

tags.tag_name,

COUNT(*) AS total

FROM photo_tags

JOIN tags

ON photo_tags.tag_id=tags.id
```

GROUP BY tags.id

ORDER BY total DESC

LIMIT 5;



Output

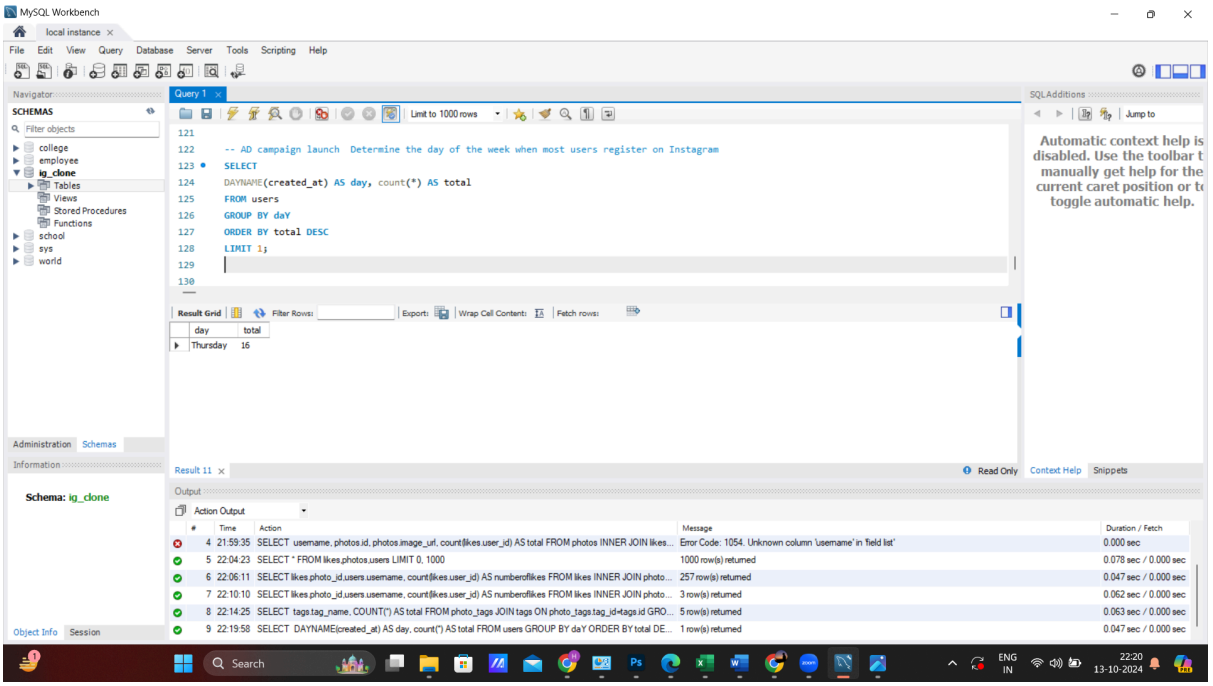
smile	59
beach	42
party	39
fun	38
concert	24

These are the top five commonly used tags with numbers.

5. AD campaign launch: The team wants to know the best day of the week to launch ads.

Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign

-- AD campaign launch Determine the day of the week when most users register on Instagram



SELECT

DAYNAME(created_at) AS day, count(*) AS total

FROM users

GROUP BY daY

ORDER BY total DESC

LIMIT 1;

Output

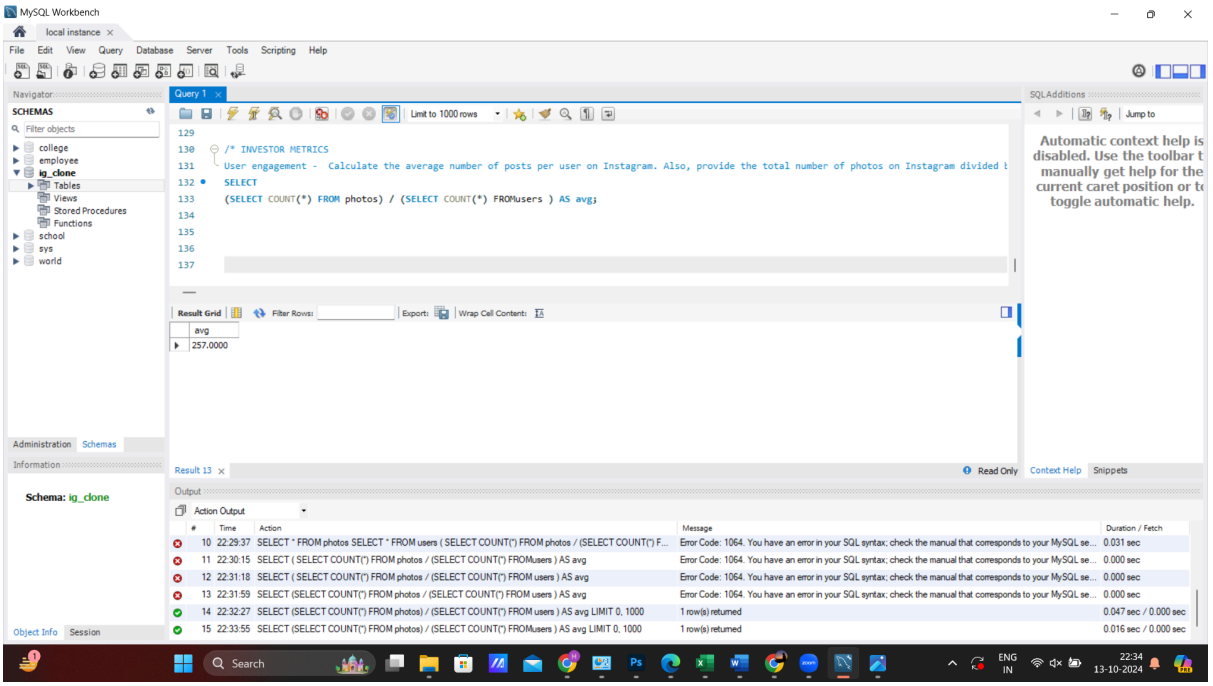
Thursday 16

Thursday is the day when most user (16) users registered.

Investor Metrics

1.User engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.



SELECT

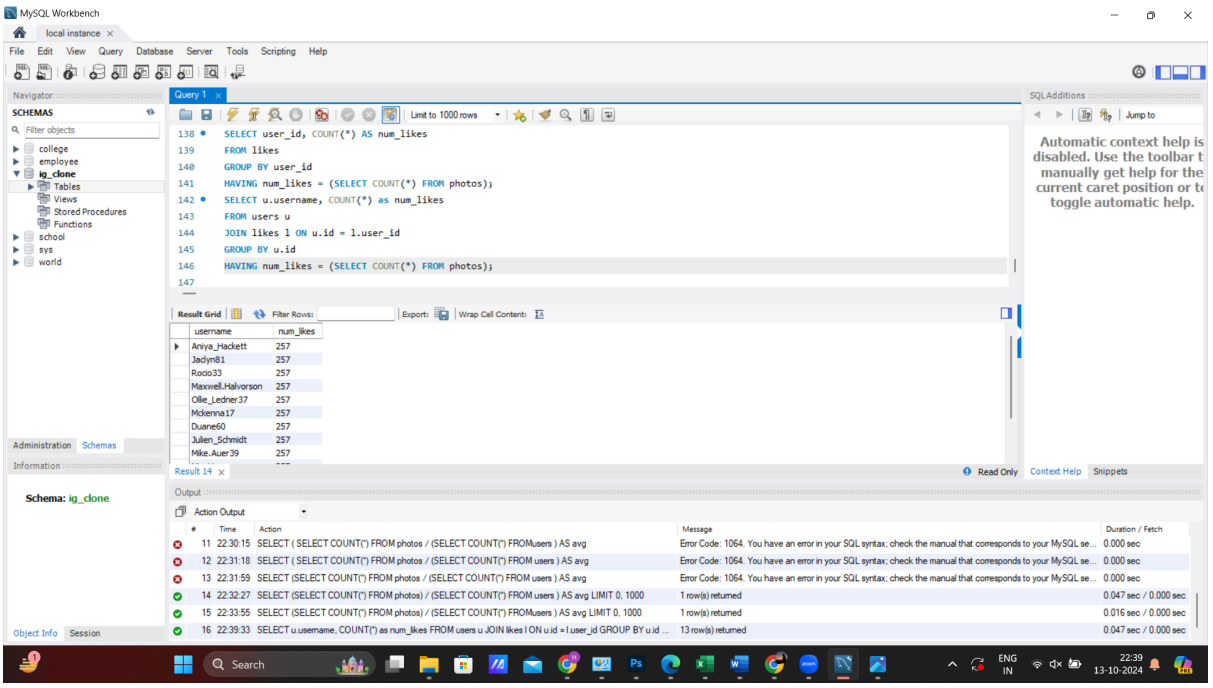
(SELECT COUNT(*) FROM photos) / (SELECT COUNT(*) FROM users) AS avg;

Output 257

The average number of posts per user on Instagram is 257.

2.BOT and FAKE accounts: Investors want to know if the platform is crowded with fake and dummy accounts.

Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.



SELECT user_id, COUNT(*) AS num_likes

FROM likes

GROUP BY user_id

HAVING num_likes = (SELECT COUNT(*) FROM photos);

SELECT u.username, COUNT(*) as num_likes

FROM users u

JOIN likes l ON u.id = l.user_id

GROUP BY u.id

HAVING num_likes = (SELECT COUNT(*) FROM photos);

Output.

Aniya_Hackett	257
Jaclyn81	257
Rocio33	257
Maxwell.Halvorson	257
Ollie_Ledner37	257
Mckenna17	257
Duane60	257
Julien_Schmidt	257
Mike.Auer39	257
Nia_Haag	257
Leslie67	257
Janelle.Nikolaus81	257
Bethany20	257

These are the fake bot accounts that has liked every single post that's not humanly possible.

Insights:

1. Oldest Users: The analysis identified the five oldest users, indicating potential candidates for loyalty rewards. The oldest user in the dataset was created on May 6, 2016.
2. Inactive Users: A significant number of users were found to have never posted, which highlighted an opportunity for re-engagement efforts.
3. Contest Winner: A clear contest winner was determined based on likes, with the user "Zack_Kemmer93" having 48 likes on their post.
4. Top Hashtags: The most popular hashtags were found to be "smile," "beach," "party," "fun," and "concert," guiding strategic hashtag use for partnerships.

5. Best Day for Ads: The analysis showed that Thursday was the day with the highest number of user registrations, suggesting it as an optimal day for launching ad campaigns.
6. Average Posts Per User: Calculated to be 257, showing the level of engagement per user.
7. Bot Detection: Multiple users were flagged as potential bots for liking every post in the database, an unlikely behavior for real users.

Results

Actionable Insights for Marketing: The project enabled the marketing team to target loyal and inactive users effectively and provided crucial data on the best times to run campaigns.

Strategic Content Creation: The data on popular hashtags helped refine content strategies for greater user reach.

Investor Confidence: By providing clear engagement metrics and identifying potential bot accounts, the project supported investor relations by showing active management of user authenticity and platform health.

Platform Improvement Suggestions: Highlighted areas for improvement in user engagement strategies and trust-building by addressing suspicious activity.

This project effectively utilized data analysis to enhance decision-making for marketing, operations, and investor communication within an Instagram-like platform.